ArtSim: Improved estimation of current impact for recent articles

Serafeim Chatzopoulos, Thanasis Vergoulis, Ilias Kanellos, Theodore Dalamagas, Christos Tryfonopoulos

AIMinScience 2020
1st International Workshop on Assessing Impact and Merit in Science

Problem definition (1/2)

- Plethora of available scientific articles
- *Scientific impact* can be defined in many ways [1]
- Variety of measures aiming to quantify scientific impact
- Practical applications
    - Keyword search results
    - Literature exploration
    - Articles comparison

[1] Bollen, J., Van de Sompel, H., Hagberg, A., Chute, R.: A principal component analysis of 39 scientific impact measures. PloS one (2009)

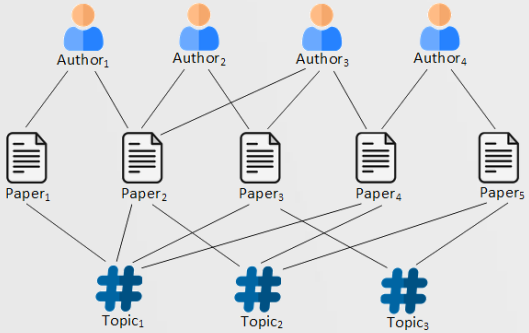
Problem definition (2/2)

- Estimate *popularity* or *hype* of an article i.e. its current impact
- Most popularity measures rely on citation history
- Recent articles have limited citation history
    - Their impact estimation is prone to inaccuracies
    - Existing measures fail to provide correct estimations

Our approach (1/3)

- ArtSim: a novel approach to assess article popularity
- Similar articles share similar popularity dynamics
- ArtSim considers similar articles with more complete citation history
- Scholarly Knowledge Graphs are used to compute article similarity
- Article similarity based on author lists and topics

Our approach (2/3)

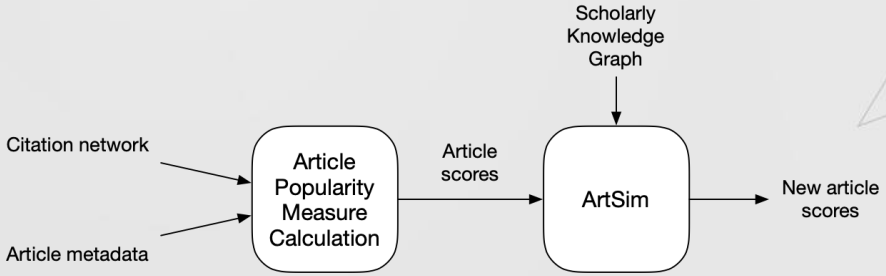


- Knowledge Graphs encode rich domain-specific information
- A *metapath* is a path in the network schema
  - APTPA relates authors that have published papers in the same topics
- ArtSim calculates *metapath-based similarity* between papers [2]

[2] Xiong, Y., Zhu, Y., Yu, P.S.: Top-k similarity join in heterogeneous information networks. IEEE Transactions on Knowledge and Data Engineering (2015)

Our approach (3/3)

- ArtSim can be applied on top of any popularity measure
- ArtSim transforms scores of recent papers using article similarity
- $S(p) = \begin{cases} a * S_{PAP}(p) + \beta * S_{PTP}(p) + \gamma * S_{initial}(p), & p.year \geq t_c - y \\ S_{initial}(p) & , \quad otherwise \end{cases}$
- $\alpha, \beta, \gamma \in [0, 1]$ and $\alpha + \beta + \gamma = 1$



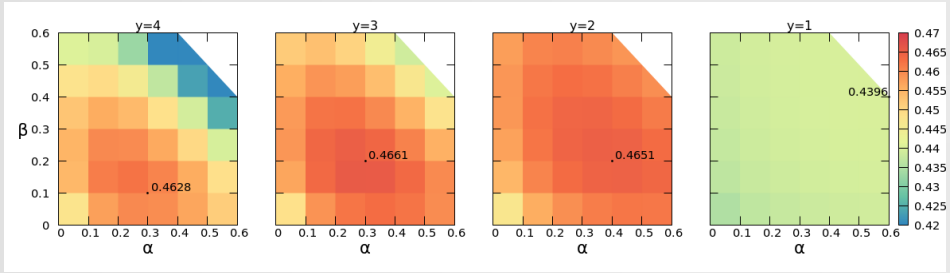
Evaluation (1/5)

- Datasets
  - DBLP Scholarly Knowledge Graph
  - DBLP Article Similarities (https://zenodo.org/record/3778916)
- Ground truth: total rankings of papers based on their short term future citations [3]
- Measures: Kendall's τ and nDCG@k
- Best performing popularity measures [3]
  - Retained Adjacency Matrix (RAM)
  - Effective Contagion Matrix (ECM)
  - CiteRank (CR)
  - FutureRank (FR)

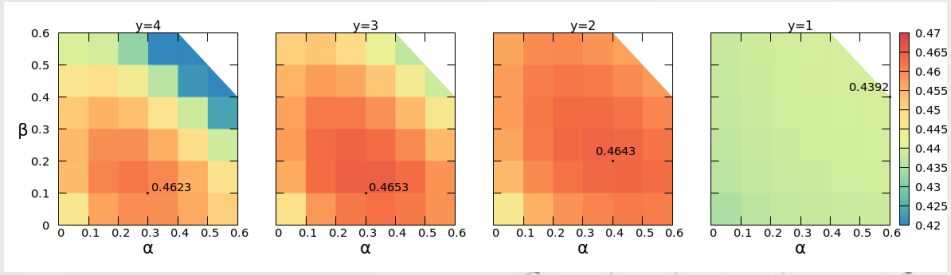
[3] Kanellos, I., Vergoulis, T., Sacharidis, D., Dalamagas, T., Vassiliou, Y.: Impactbased ranking of scientific publications: A survey and experimental evaluation. IEEE Transactions on Knowledge and Data Engineering (2019).

Evaluation (2/5)

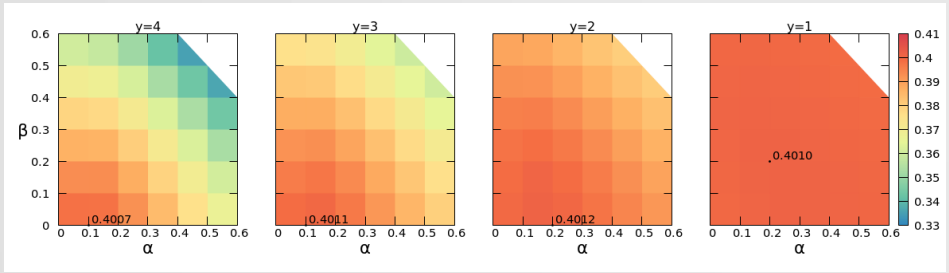
Effectiveness of our approach for different parameters in terms of Kendall's τ correlation



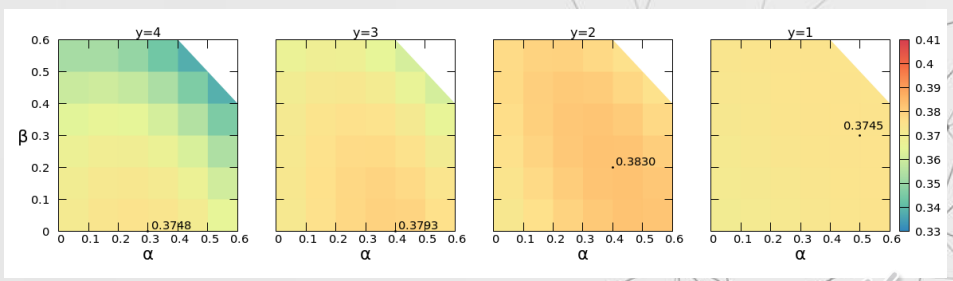
ECM



RAM



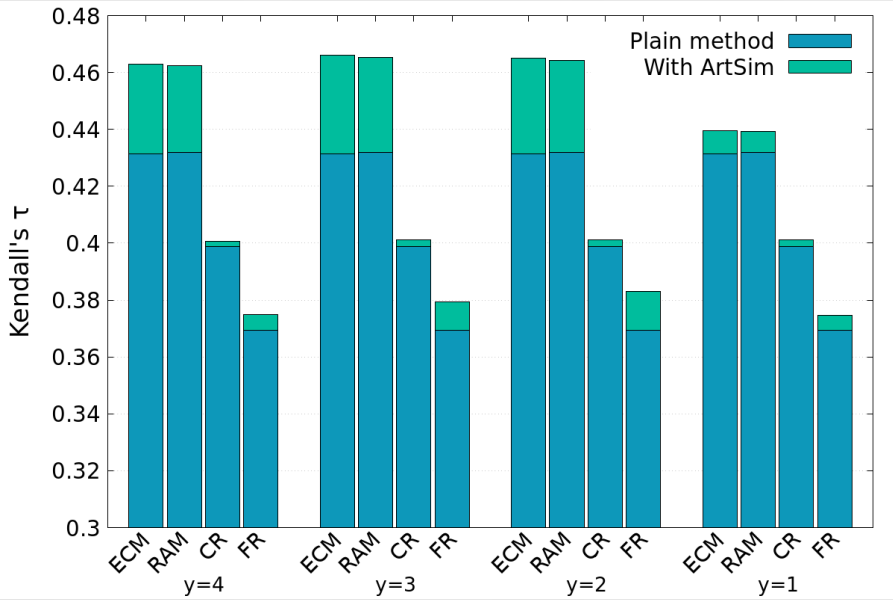
CR



FR

Evaluation (3/5)

Effectiveness of our approach in terms of Kendall's τ correlation for best parameterization per year



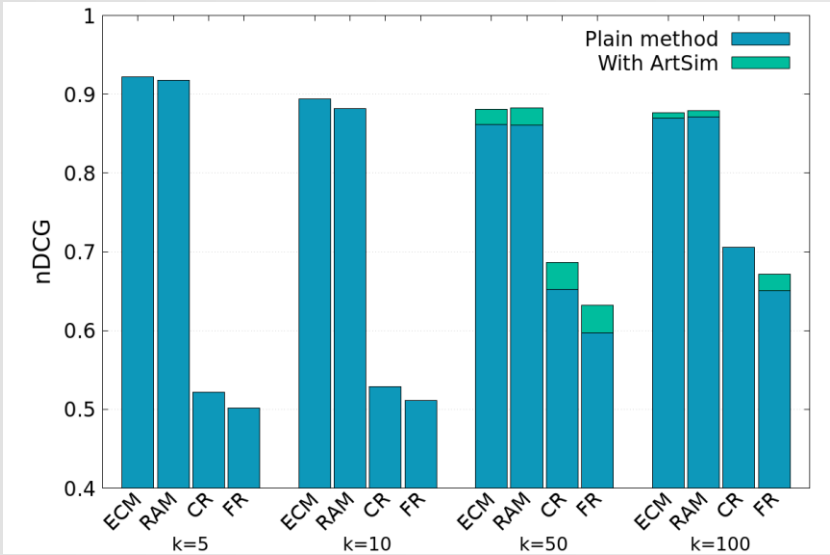
Evaluation (4/5)

Effectiveness of our approach for y = 3 in terms of nDCG@k

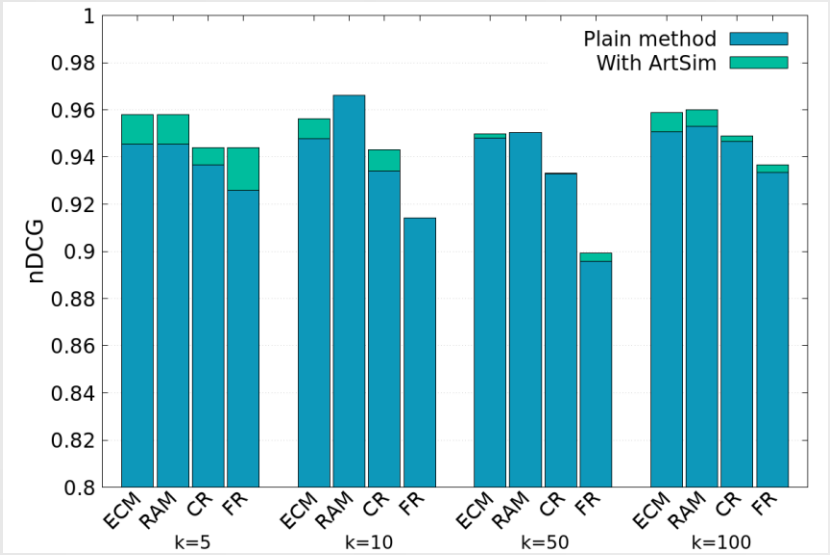
	Small values of k			Large values of k		
	5	**50**	**500**	**400,000**	**500,000**	**600,000**
ECM	0.8323	0.8634	0.8953	0.8780	0.8833	0.8884
ArtSim-ECM	0.8323	0.8634	0.8953	**0.8837**	**0.8912**	**0.9003**
RAM	0.8588	0.8521	0.8943	0.8774	0.8842	0.8881
ArtSim-RAM	0.8588	0.8521	0.8943	**0.8836**	**0.8904**	**0.9008**
CR	0.3530	0.5263	0.6060	0.7904	0.8149	0.8272
ArtSim-CR	0.3530	0.5263	0.6060	**0.7983**	**0.8199**	**0.8307**
FR	0.3403	0.5018	0.5526	0.7586	0.7934	0.8101
ArtSim-FR	0.3403	0.5018	0.5526	**0.7731**	**0.7961**	**0.8152**

Evaluation (5/5)

Effectiveness of our approach for y = 3 in terms of nDCG for different keyword search scenarios with y = 3 and varying k



Expert Finding



Recommender Systems

Summarizing

- ArtSim can be applied on top of existing popularity measures to increase the accuracy of their results
- Popularity of papers in their cold start period can be better estimated based on the characteristics of other, similar papers
- Experimental evaluation showcases the effectiveness of ArtSim

Thank you for your attention!

schatz@athenarc.gr

European Union
European Regional Development Fund

HELLENIC REPUBLIC
MINISTRY OF ECONOMY & DEVELOPMENT
SPECIAL SECRETARY FOR ERDF & CF
MANAGING AUTHORITY OF EPAnEK

EPAnEK 2014-2020
OPERATIONAL PROGRAMME
COMPETITIVENESS
ENTREPRENEURSHIP
INNOVATION

ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη
Partnership Agreement 2014 - 2020

Co-financed by Greece and the European Union